·医学循证·

基于行政管理数据识别慢性病发病病例背景下确定洗脱期时长最佳策略的系统综述

杨文怡，王敬鑫，艾丽梅，万霞[*]

【摘要】 **背景** 使用行政管理数据时，确立清晰、适当的慢性病洗脱期时长是正确确定反复就医的慢性病患者发病时点、确定新发病例的基础。**目的** 通过系统文献回顾，综述确定洗脱期时长的方法，以期为我国研究者后续使用行政管理数据识别慢性病新发病例时确认洗脱期长短、正确识别新发病例提供思路。**方法** 于2021年10月，系统检索PubMed、Web of Science、EmBase、中国知网、维普中文科技期刊全文数据库、万方知识服务平台，获取有关利用行政管理数据探究慢性病发病、患病情况的文献，检索时限均为建库至2022-10-01。由两名研究者独立筛选文献并提取相关信息，并采用定性研究报告评价标准（SRQR）评价文献方法学质量后，使用描述性分析法总结洗脱期时长的确定方法。**结果** 共纳入26篇文献，纳入文献的SRQR评分均≥15分，方法学质量较好。文献所使用的数据主要来自加拿大、美国、澳大利亚等行政管理数据完整、丰富的国家（地区），聚焦的疾病包括糖尿病、肿瘤、精神分裂症等多种慢性病。研究指出，设定合适的洗脱期时长是准确识别发病病例的基础。目前，文献中确定洗脱期时长的方法主要包括直接限定法、一致性检验法和逆向生存函数法三大类，其中最常用的方法是直接限定法，逆向生存曲线法的使用率相对较低。**结论** 直接限定法、一致性检验法和逆向生存函数法均有相应的优势和局限性，方法的选择标准、判断标准和稳定性有待进一步探究。

【关键词】 慢性病；行政管理数据；医疗保险数据；洗脱期；发病率；患病率

【中图分类号】 R 195 【文献标识码】 A DOI：10.12114/j.issn.1007-9572.2023.0005

【引用本文】 杨文怡，王敬鑫，艾丽梅，等.基于行政管理数据识别慢性病发病病例背景下确定洗脱期时长最佳策略的系统综述［J］.中国全科医学，2023.［Epub ahead of print］DOI：10.12114/j.issn.1007-9572.2023.0005.［www.chinagp.net］

YANG W Y，WANG J X，AI L M，et al. Optimal strategies for determining the duration of washout period in the context of identifying chronic disease onset casesbased on administrative data: a systematic review［J］. Chinese General Practice，2023.［Epub ahead of print］

Optimal Strategies for Determining the Duration of Washout Period in the Context of Identifying Chronic Disease Onset Cases Based on Administrative Data：a Systematic Review *YANG Wenyi，WANG Jingxin，AI Limei，ZHAO Jian，WAN Xia*

Institute of Basic Medical Sciences，Chinese Academy of Medical Science & School of Basic Medicine，Peking Union Medical College，Beijing 100005，China

[*]*Corresponding author：Wan Xia，Doctoral supervisor/Professor；E-mail：xiawan@ibms.pumc.edu.cn*

【**Abstract**】 **Background** When using administrative data，the establishment of clear and appropriate duration of chronic disease washout period is the basis for correctly identifying the point of onset of chronic disease in patients with recurrent visits and identifying new cases. **Objective** To review the methods for determining the duration of washout period by a systematic review of literature，in order to provide ideas for Chinese researchers to confirm the duration of washout period and correctly identify new cases with subsequent use of administrative data to identify new cases of chronic diseases. **Methods** In October 2021，PubMed，Web of Science，EmBase，CNKI，CQVIP and Wanfang Knowledge Service Platform were systematically searched for literature on the use of administrative data to explore the incidence and prevalence of chronic diseases from inception to 2022-10-01. Two researchers independently screened literatures and extracted relevant information，evaluated the quality of literatures by using the Standards for Reporting Qualitative Research（SRQR）scale，and summarized the determination method

基金项目：美国中华医学基金会项目（15-208）

100005 北京市，中国医学科学院基础医学研究所　北京协和医学院基础学院流行病与卫生统计学系

[*]通信作者：万霞，博士生导师/研究员；E-mail：xiawan@ibms.pumc.edu.cn

本文数字出版日期：2023-07-13

of the duration of washout period. **Results** A total of 26 papers were included, and the SRQR scores of the included literature were all ≥ 15, indicating good methodological quality. The data used in the literature were mainly from Canada, the United States, Australia and other countries (regions) with complete and abundant administrative data, and focused on a variety of chronic diseases, including diabetes, tumor, schizophrenia and other chronic diseases. The study pointed out that setting an appropriate washout period duration is the basis for accurate identification of onset cases. The methods used to determine the duration of washout period in literature were mainly divided into three major categories, including direct restriction method, consistency test method and retrograde survival function method. Among them, the most commonly used method was direct restriction method, while the retrograde survival function method had a relatively low usage rate. **Conclusion** Direct restriction method, consistency test method and retrograde survival function method all have corresponding advantages and limitations. The selection, judgment criteria and stability of the method need to be further explored.

【**Key words**】 Chronic disease; Administrative data; Insurance data; Washout period; Incidence; Prevalence

为了有效预防和控制糖尿病、肿瘤、慢性阻塞性肺疾病等多种慢性病，需要准确估计疾病的患病率和发病率。目前，已有一些研究者通过开展大型队列、调查研究对疾病的患病、发病情况展开调查，但这种研究方式具有耗时长、耗资大的特点。因此，可考虑应用医疗保险报销记录、健康监测数据等行政管理数据对慢性病的流行病学特征进行分析。与患者从出生到死亡的全生命周期的医疗记录不同，医疗保险数据多为有限年份下的记录，且慢性病患者通常会反复就医，从这些被截断的、高度重复的记录中有效识别出新发病例有一定难度。为解决这一问题，研究者一般以回顾期（look-back period），即洗脱期（washout period）为基础确定新发病例。具体方法为：若在目标检索年限内患者出现在某种慢性病的登记记录中，即可认为其为疑似新发病例，以目标检索年限为基础，向前回顾，在限定的洗脱期内没有与该病相关的记录即可确定其为新发病例。目前，对于洗脱期的长短尚没有统一结论[1-2]。洗脱期太短会造成发病率被高估，太长则会导致数据利用不足[3]。一些研究者认为不同疾病有不同的发展轨迹和特征[4-6]，在使用不同类型数据确定不同疾病新发病例时，均应采用合适的方法确定最佳洗脱期时长。本研究系统综述了确定洗脱期时长的方法，以期为我国研究者后续使用行政管理数据识别慢性病新发病例时确认洗脱期长短、正确识别新发病例提供思路。

1 资料与方法

1.1 文献纳入与排除标准 文献纳入标准：（1）所使用的数据为行政管理数据，包括医疗保险数据、疾病登记注册数据、医院登记数据等；（2）所聚焦的疾病为慢性病，同一患者因同种疾病产生多次记录；（3）研究目的为探究疾病患病、发病情况，聚焦疾病流行病学特征；（4）涉及洗脱期。文献排除标准：（1）非中、英文文献；（2）所使用的数据为非官方数据，如通过队列研究、病例对照研究获得的调查数据等；（3）非原始研究，如评论、系统综述等；（4）会议论文或以摘要形式发表的研究成果；（5）无法获取原文的文献；（6）重复发表的文献；（7）所探讨的内容与本研究关注的内容不相关，包括：仅利用行政管理数据评估疾病负担、开展队列 / 病例对照研究等。

1.2 文献检索策略 于2021年10月，系统检索3个英文数据库（PubMed、Web of Science、EmBase）和3个中文数据库（中国知网、维普中文科技期刊全文数据库、万方知识服务平台），获取有关利用行政管理数据探究慢性病发病、患病情况的文献，检索时限均为建库至2022-10-01。中文检索词包括：医疗保险数据、官方数据、行政数据、医院数据、患病、发病，洗脱期、回顾期、窗口期。英文检索词包括：administrative data、insurance data、hospital data、Medicare，prevalence、incidence，look-back period、wash-out time、clearance time、disease-free time、observation time。以PubMed为例，具体检索策略如下：（incidence［Title/Abstract］OR prevalence［Title/Abstract］）AND（insurance data［Title/Abstract］OR administrative data［Title/Abstract］OR hospital data［Title/Abstract］OR Medicare［Title/Abstract］）AND（look-back period［Title/Abstract］OR wash-out time［Title/Abstract］OR clearance time［Title/Abstract］OR disease-free time［Title/Abstract］OR observation time［Title/Abstract］）。

1.3 文献筛选与资料提取 由2名研究人员独立筛选文献、提取资料，并交叉核对结果，如有分歧，通过与第3名研究者讨论解决。文献筛选时，首先阅读文题和摘要，在排除明显不相关的文献后，进一步通读全文根据纳入和排除标准确定纳入分析的文献。提取的资料包括：（1）文献基本信息，作者、发表年份、聚焦疾病；（2）使用的数据库的基本信息，国家（地区）、数据库名称、数据库类型、使用数据的年份跨度、涵盖人群（数）；（3）研究设计，最佳洗脱期时长的判断方法及标准、设定的洗脱期时长。

1.4 纳入文献方法学质量评价 由2名研究人员独立

采用美国加利福利亚大学提出的定性研究报告评价标准（Standards for Reporting Qualitative Research，SRQR）[7] 对文献方法学质量进行评价，如有分歧，通过与第 3 名研究者讨论解决。SRQR 共有 21 个条目，研究者对每个条目做出“是”“否”“不清楚”的判断，选择“是”计 1 分，选择“否”“不清楚”计 0 分，将各条目得分相加即得总分，得分范围为 0~21 分，最终纳入 SRQR 评分≥ 15 分的文献。

1.5　统计学方法　使用描述性分析法总结洗脱期时长确定方法。

2　结果

2.1　文献筛选流程及结果　通过检索各数据库获得中文文献 21 篇、英文文献 1 203 篇。经过初筛、复筛，获得基于行政管理数据评估慢性病患病 / 发病情况等流行病学特征的文献 54 篇，排除其中未涉及洗脱期的文献 28 篇，最终纳入 26 篇[2-6，8-28] 文献。文献筛选流程具体见图 1。

2.2　纳入文献基本信息及使用的数据库的基本信息　纳入的 26 篇文献均为英文文献；所使用的数据主要来自加拿大、美国、澳大利亚等行政管理数据完整、丰富的国家（地区）；所聚焦的疾病包括糖尿病、肿瘤、精神分裂症等多种慢性病，其中 11 篇[2，4，6，8，14，17，19，22-23，25，28] 文献聚焦心血管疾病，5 篇[6，8，10，13，19] 文献聚焦糖尿病；研究目的主要为确定多种慢性病的发病、患病情况；12 篇[3，6，9-11，13，16，19-20，24，26-27] 文献所使用的数据为医疗保险数据，8 篇[2，4，8，12，22-23，25，28] 文献所使用的数据为医院患者注册登记数据，其余文献所使用的数据为经注册的居民健康项目数据[14-15] 或经注册的疾病调查项目数据[5，17-18，21]；所使用的数据年份跨度较大，其中仅有两篇[15，19] 文献所使用的数据年份跨度 <5 年，14 篇[2，5，8，10-12，14，16，17，21，24-26，28] 文献所使用的数据年份跨度≥ 10 年。纳入文献基本信息及使用的数据库的基本信息见表 1。

2.3　纳入文献的方法学质量评价结果　纳入文献的 SRQRQ 评分均≥ 15 分（受限于篇幅，本研究未通过单独列表呈现纳入研究的方法学质量评价具体结果），见表 1。

2.4　最佳洗脱期时长确定方法　通过系统文献回顾，发现目前确定洗脱期时长的方法主要包括直接限定法、一致性检验法和逆向生存函数法三大类。纳入的 26 篇文献中，有 15 篇[3，13-15，18-28] 基于专家建议、临床经验、文献回顾或因受限于数据，直接确定洗脱期时长；10 篇文献[2，4-6，9-12，16-17] 使用一致性检验的相关指标，如 Kappa 值、阳性预测值、过高估计率等评估不同洗脱期时长对发病率估计的影响、确定洗脱期时长；3 篇[8-10] 使用逆向生存函数法确定最佳洗脱期时长。

2.4.1　一致性检验法　10 篇[2，4-6，9-12，16-17] 文献所采用的一致性检验相关指标包括 Kappa 值、阳性预测值、过高估计率、阴性预测值、灵敏度、准确度。所有文献均将基于数据实际所能达到的最长洗脱期期限作为洗脱期时长“金标准”，通过以上指标定量比较不同洗脱期时长下的发病人数（Incident patient defined by different clearance period，IDD）与基于洗脱期时长“金标准”确定的发病人数（Incident patient defined by golden criterion，IDG）之间的差异性。6 篇[2，4-6，16-17] 文献采用过高估计率〔过高估计率 =（IDD-IDG）/IDG × 100%〕作为一致性评价指标，3 篇[9-11] 文献所采用的一致性评价指标涉及 Kappa 值，3 篇[9，11-12] 文献所采用的一致性评价指标涉及阳性预测值，1 篇[12] 文献所采用的一致性评价指标涉及灵敏度、特异度和阴性预测值。目前，仅有研究者基于 Kappa 值提出了一致性的判断标准。按照 BYRT 等[29] 提出的判断标准：当 kappa 值为 -1.00~<0 时，可认为“一致性缺乏”；当 kappa 值为 0~0.20 时，可认为“一致性不佳”；当 kappa 值为 >0.20~0.40 时，可认为“低一致性”；当 kappa 值为 >0.40~0.60 时，可认为“一致性一般”；当 kappa 值为 >0.60~0.80 时，可认为“一致性较好”；

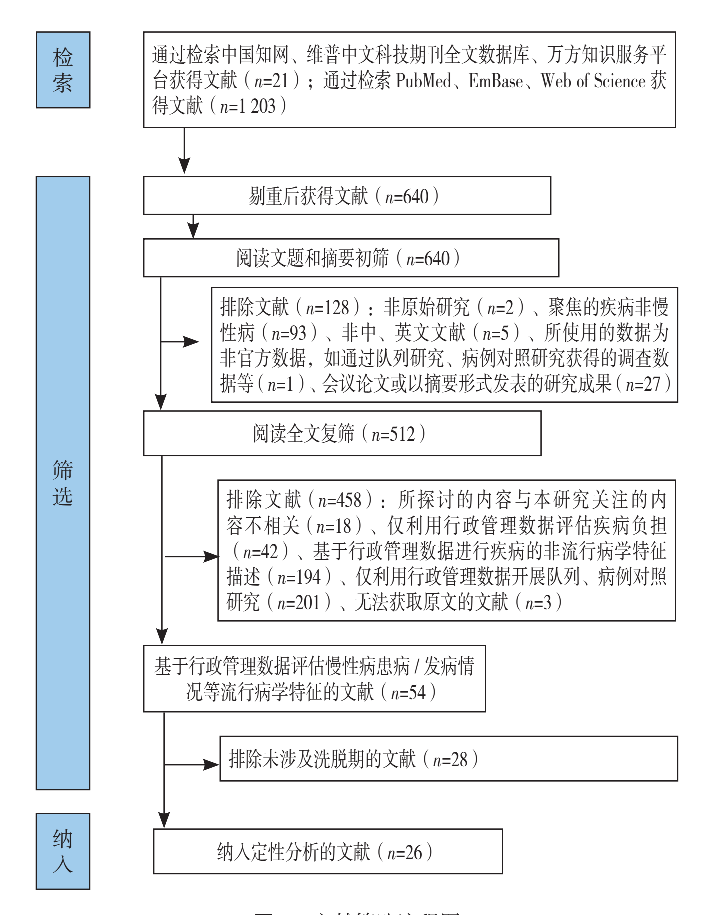


图 1　文献筛选流程图
Figure 1　Literature screening flow chart

表 1 纳入文献基本信息及使用的数据库的基本信息

Table 1 Basic information of the included literature and the used database

作者	发表年份（年）	聚焦疾病	SRQR 评分（分）	国家（地区）	数据库名称	数据库类型	使用数据的年份跨度（年）	涵盖人群（数）
ØDEGAARD 等[2]	2022	心力衰竭	17	挪威	Norwegian Patient Registry	医院患者注册登记数据	2008—2018	挪威全体人群
BRAMELD 等[8]	2003	糖尿病、心肌梗死	16	美国	Patient Registry	医院患者注册登记数据	1980—1995	—
BEAUDET 等[9]	2014	下背痛	17	加拿大	R é gie de l'Assurance Maladie du Qu é bec（RAMQ）	医疗保险数据	2000—2007	魁北克省 50% 以上人口
ASGHARI 等[10]	2009	糖尿病	16	加拿大	RAMQ	医疗保险数据	1992—2002	568 625 5
VANASSE 等[11]	2012	精神分裂症	16	加拿大	RAMQ	医疗保险数据	1996—2006	魁北克省 50% 以上人口
BENCHIMOL 等[12]	2014	慢性肠道疾病	16	加拿大	Canadian Institute for Health Information's Discharge Abstract Database（CIHI-DAD）	医院患者注册登记数据	2002—2012	—
LIN 等[13]	2014	糖尿病视网膜病变	17	中国台湾省	Health Insurance Database of Taiwan	医疗保险数据	2005—2011	台湾省 99% 人口
SULO 等[14]	2015	心肌梗死	15	挪威	Norway project	经注册的居民健康项目数据	1994—2009	—
CZWIKLA 等[3]	2017	乳腺癌、前列腺癌、结直肠癌	16	德国	Statutory Health Insurance	医疗保险数据	2006—2014	德国 90% 人口
RASSEN 等[15]	2019	慢性阻塞性肺疾病、银屑病、精神分裂症、囊性纤维病	16	美国	electronic health records	经注册的居民健康项目数据	2011—2014	美国大部分人口
ABBAS 等[6]	2012	糖尿病、结直肠癌、心力衰竭	16	德国	Statutory Health Insurance	医疗保险数据	2000—2008	德国 90% 人口
STEIN 等[16]	2013	青光眼、黄斑变性	17	美国	Clinformatics Data Mart database	医疗保险数据	2001—2011	2 457
ROSENLUND 等[4]	2020	24 种慢性病	17	瑞典	National Patient Register	医院患者注册登记数据	2005—2013	7 943 807
SCHMEDT 等[5]	2017	多发性硬化	17	德国	German Pharmaco epide miological Research Database（GePaRD）	经注册的疾病调查项目数据	2004—2013	4 175 877
CAMPLAIN 等[17]	2017	心力衰竭	18	美国	Atherosclerosis Risk in Communities（ARIC）	经注册的疾病调查项目数据	2000—2012	15 744
OHLMEIER 等[18]	2019	亨廷顿病	17	德国	Applied Health Research Berlin（InGef）Research Database	经注册的疾病调查项目数据	2010—2016	约 670 万人
KOO 等[19]	2014	糖尿病、动脉粥样硬化	17	韩国	National Health Insurance（NHI）Database	医疗保险数据	2007—2011	韩国 97% 人口
JÜRISSON 等[20]	2015	髋骨骨折	17	爱沙尼亚	Estonian Health Insurance Fund	医疗保险数据	2005—2012	爱沙尼亚 96% 以上人口
BUKHBINDER 等[21]	2022	阿尔茨海默病	17	美国	Optum Clinformatics Data Mart（CDM）	经注册的疾病调查项目数据	2009—2019	—
ZHOU 等[22]	2022	血栓形成	17	加拿大	CIHI-DAD	医院患者注册登记数据	2011—2017	—
KATZENELLENBOGEN 等[23]	2019	风湿性心脏病	17	澳大利亚	acute rheumatic fever register data	医院患者注册登记数据	2000—2017	—
OU 等[24]	2012	青光眼、阿尔茨海默病、其他类型痴呆	18	美国	Medicare 5% claims data	医疗保险数据	1994—2007	484 542
CLISSOLD 等[25]	2017	脑卒中	17	加拿大	Victorian Admitted Episodes Dataset	医院患者注册登记数据	1997—2008	4 932 422
AHN 等[26]	2021	风湿性疾病	17	韩国	NHI Database	医疗保险数据	2007—2018	韩国 97% 人口
ANDERSON 等[27]	2012	精神分裂症	18	加拿大	RAMQ	医疗保险数据	2000—2006	魁北克省 50% 以上人口
KATZENELLENBOGEN 等[28]	2010	心肌梗死	20	澳大利亚	Hospital Morbidity Data Collection（HMDC）、Mortality Data Set	医院患者注册登记数据	1980—2004	—

注：—表示未提及。

当 kappa 值为 >0.80~0.90 时，可认为“一致性好”；当 kappa 值为 >0.90~1.00 时，可认为“一致性极好”。当 kappa 值为 >0.90~1.00 时，所选洗脱期时长下的发病人数与基于洗脱期时长“金标准”确定的发病人数之间的一致性极好，所选洗脱期时长即为最佳洗脱期时长。

2.4.2 逆向生存函数法 一般通过 Kaplan-Meier 法或寿命表法实现估计。在该方法中，终点事件为患者在洗脱期内被记录，生存被定义为在给定的洗脱期内没有捕捉到该患者的记录（此时患者可被认为是新发病例）。具体来说，如计算 j 年某慢性病 i 的发病人数：a 代表某患者 j 年第一次因疾病 i 被记录的日期，b 代表该患者洗脱期内离 a 最近的一次因疾病 i 被记录的日期，c 代表洗脱期的开始日期。如果该患者在给定的洗脱期内被记录，该患者的生存时间为 $l_x=a-b$；如果该患者在给定的洗脱期内没有被记录，则该患者被定义为“删失”，即“生存”，删失时间为 $l_c=a-c$。据此建立生存函数，使用 Kaplan-Meier 法或寿命表法计算在给定洗脱期 T 下，研究对象的生存概率〔$S(t)$〕，并绘制生存曲线。同时，BRAMELD 等[8] 和 BEAUDET 等[9] 在研究中运用风险函数〔$h(t)$〕计算了生存时间已达到 t 的个体在接下来的 Δt 时间内发生终点事件的瞬时风险，并绘制了风险函数曲线。研究认为，若存在某一时点 tf 使得 $h(t_f)$ 趋近于零、$S(t_f)$ 趋近于定值，该时点即代表最佳洗脱期时长。在实际研究中，BRAMELD 等[8] 认为，当 $h(t_f)<0.000\ 01$ 时，即可认为 $h(t_f)$ 趋近于 0，而 BEAUDET 等[9] 则未给出具体标准。ASGHARI 等[10] 未通过风险函数计算生存时间已达到 t 的个体在接下来的 Δ_t 时间内发生终点事件的瞬时风险，其认为当在某一时点 t_f 下研究对象的生存概率不再变动，即趋于稳定时，该时点即代表最佳洗脱期时长。

2.5 洗脱期时长 纳入文献所设定的洗脱期时长为 0.5~15.0 年。当不同文献聚焦的疾病不同时，设定的洗脱期时长之间有明显差异；但当不同文献聚焦的疾病相同时，设定的洗脱期时长之间的差异也较大，如 5 篇[6, 8, 10, 13, 19] 聚焦糖尿病的文献所设定的洗脱期时长，最长为 13 年，最短为 1 年。纳入文献洗脱期时长确定方法及设定的洗脱期时长见表 2。

3 讨论

本次研究系统梳理了使用行政管理数据识别慢性病发病病例时确定最佳洗脱期时长的方法。目前，半数左右〔48.1%（26/54）〕基于行政管理数据评估慢性病患病 / 发病情况等流行病学特征的文献在识别慢性病发病病例时，设定了洗脱期，但对于洗脱期的时长目前尚没有统一结论。纳入的 26 篇文献所设定的洗脱期时长为 0.5~15.0 年。确定洗脱期时长的方法主要包括直接限定法、一致性检验法、逆向生存函数法 3 种，其中最常用的方法是直接限定法，逆向生存曲线法的使用率相对较低。

本次研究发现，半数以上〔57.7%（15/26）〕被纳入文献基于直接限定法给定洗脱期时长，在该方法下，研究者根据专家建议、文献回顾结果等直接限定洗脱期时长。直接限定法虽具有便捷、易于操作的特点，但因缺乏量化的数据支持，一定程度上使限定的洗脱期时长不具有说服力。部分研究者也在研究局限性部分指出，直接限定洗脱期时长可能会导致发病率估计结果不准确[18, 26]；2 篇[14, 23] 文献选择直接以（可获得数据的最大年份 - 可获得数据的最小年份）为洗脱期时长，但此方法下研究者只能探讨可获得数据最大年份慢性病的发病情况，数据无法得到有效利用。部分文献〔38.5%（10/26）〕基于一致性检验的相关指标确定最佳洗脱期时长[2, 4-6, 9-12, 16-17]，但不同的文献所采用的一致性检验指标不同，目前被使用较多的一致性检验指标为 Kappa 值和过高估计率，其中 Kappa 值的判定标准为 BYRT 等[29] 1993 年提出的标准。多数研究者选择 Kappa 值 >0.80（“一致性好”/“一致性极好”）作为确定洗脱期时长的标准[9-11]，但有研究认为在识别高患病率疾病（如糖尿病、心血管疾病）发病病例时，高 Kappa 值可能与疾病患病人数多有关，Kappa 值的稳定性有待进一步探究[10]。目前，尚缺乏基于除 Kappa 值外的其他一致性评价指标的洗脱期时长确定标准，如 ABBAS 等[6] 以过高估计率 10% 作为临界标准，而在 ROSENLUND[4] 等的研究中这一指标被设定为 20%；对于阳性预测值，有研究者选择 80% 作为界值[11]；BENCHIMOL 等[12] 则选择准确度达到 90% 作为评判标准。仅有少部分研究〔11.5%（3/26）〕使用逆向生存函数法确定最佳洗脱期时长[8-10]。在 BRAMELD 等[8] 于 2003 首次提出采用逆向生存函数法确定洗脱期时长后，该方法被陆续使用[9-10]。逆向生存函数法强调通过绘制生存曲线和风险函数曲线，并依据生存概率和风险概率确定最佳洗脱期时长。研究者认为该方法可以使数据得到最大限度的利用，并能够确保研究者充分使用所有可获得的记录估计生存函数。同时，通过生存曲线和风险函数曲线可以直观观察到生存概率和风险概率的集中趋势，有效、定量地判断最佳的洗脱期时长。但该方法也存在一定的局限性，即判断标准尚未得到完全统一。如 BRAMELD 等[8] 认为，若存在某一时点 tf 使得 $h(tf)<0.000\ 01$，该时点即代表最佳洗脱期时长，按照此标准，其研究（聚焦糖尿病）中洗脱期长达 13 年。ASGHARI 等[10]（其研究亦聚焦糖尿病）并未使用这一标准，其根据研究对象的生存概率趋于稳定这一标准，确认洗脱期时长为 5 年。值得注意的是，由于对洗脱期时长的确定方法、判断标准尚未达成共识，许多研究者在探究最

表 2 纳入文献洗脱期时长确定方法及设定的洗脱期时长

Table 2 Methods for determining the optimal washout period duration and the set duration of it included in the literature

作者	设定的洗脱期时长（年）	洗脱期时长确定方法		
		直接限定法	一致性检验法	逆向生存函数法
ABBAS 等[6]	1.0~8.0		√（过高估计率）	
STEIN 等[16]	3.0~5.0		√（过高估计率）	
ROSENLUND 等[4]	0.5~9.0		√（过高估计率）	
SCHMEDT 等[5]	8.0		√（过高估计率）	
CAMPLAIN 等[17]	0.5~3		√（过高估计率）	
ØDEGAARD 等[2]	1.0~8.0		√（过高估计率）	
VANASSE 等[11]	7.0		√（Kappa 值、阳性预测值）	
BENCHIMOL 等[12]	4.0		√（灵敏度、准确度、阳性预测值、阴性预测值）	
BEAUDET 等[9]	7.0		√（阳性预测值、Kappa 值）	√
ASGHARI 等[10]	5.0		√（Kappa 值）	√
BRAMELD 等[8]	13.0			√
LIN 等[13]	3.0	√（基于文献回顾）		
OHLMEIER 等[18]	4.0	√（基于文献回顾）		
KOO 等[19]	2.0	√（基于文献回顾）		
JÜRISSON 等[20]	1.0	√		
BUKHBINDER 等[21]	6.0	√		
ZHOU 等[22]	1.0	√		
KATZENELLENBOGEN 等[23]	13.5	√（因受限制于数据）		
OU 等[24]	3.0	√（基于文献回顾）		
CLISSOLD 等[25]	5.0	√（基于文献回顾）		
AHN 等[26]	3.0	√		
ANDERSON 等[27]	4.0~6.0	√		
KATZENELLENBOGEN 等[28]	15.0	√（因受限制于数据）		
SULO 等[14]	3.0、5.0、7.0、8.0、10.0	√		
CZWIKLA 等[3]	1.0~7.0	√（因受限制于数据）		
RASSEN 等[15]	1.0~4.0	√（因受限制于数据）		

佳洗脱期时长时，选择计算多个一致性指标的数值，并结合基于逆向生存函数法得出的结果进行综合判断[8-10]。

研究纳入的文献所设定的洗脱期时长为 0.5~15.0 年，聚焦不同疾病的研究设定的洗脱期时长之间的差异较大。有研究者认为，不同疾病的发展轨迹和特征不同是导致该现象的出现重要原因[6]。本研究同时发现，不同研究即使聚焦同一疾病，如糖尿病，所设定的洗脱期时长之间的差异也较大，其原因可能是不同研究所使用数据的收集方式不同且来源人群的就医习惯存在较大差异。因此，在使用不同类型的数据识别不同疾病的发病病例时，均应选择合适的方法探索最佳洗脱期时长，在此基础上针对疾病流行病学特征进行进一步的研究，这与部分研究者的观点相一致[2-3, 16]。研究者后续在使用中国医疗保险数据、疾病登记注册数据等行政管理数据进行慢性病的流行病特征相关分析时，应考虑到不同国家（地区）在医疗保险政策、环境、人种等方面存在明显差异，不应直接应用其他国家研究者得出的洗脱期时长，应在数据的支持下，选择合适的方法确定最佳洗脱期时长，并在此基础上进行后续的分析研究。

行政管理数据作为宝贵的资源，涵盖范围广、年份跨度大、可被便捷地获取，相较于传统的调查数据有一定的优势。自 2012 年改革医疗保险制度以来，中国大力推广医疗保险政策，扩大参保人群覆盖范围[30-31]，随着改革的逐步深入，有效数据的积累量持续增加。但目前在万方知识服务平台、中国知网、维普中文科技期刊全文数据库检索到的利用医保数据、医院患者注册登记数据等行政管理数据对慢性病流行病学特征进行分析的研究较少（截至 2022-10-01，仅 21 篇），我国相关研究数量远少于行政管理数据完备、研究范式成熟的其他国家（地区），说明中国在此方面仍处于探索阶段，未来发展潜力巨大。同时，中国内地研究者在使用医疗保险数据等行政管理数据进行分析时，未重视洗脱期（在万方知识服务平台、中国知网、维普中文科技期刊全文数据库检索到的基于行政管理数据评估慢性病患病 / 发病情况等流行病学特征的文献均未涉及洗脱期），此外

也未有内地研究者探讨利用中国人群医疗保险数据等行政管理数据对疾病流行病学特征进行分析时如何确定洗脱期时长。因此，本研究系统总结的 3 种可用于确定洗脱期时长的方法，可为后续研究者基于中国医疗保险数据等行政管理数据准确识别慢性病新发病例、探究疾病流行病学特征提供相关思路和方法支持。

本研究存在一定的局限性。首先，由于行政管理数据具有特殊性，部分数据可能为内部数据、机密数据，在检索文献的过程中，一些未公开发表的文献无法被检索到，故未将其纳入研究范畴，这可能会对研究结果产生一定的影响。其次，研究者仅对中、英文数据库中的相关文献进行了检索，未检索以其他语种发表的相关文献，可能会导致检索到的方法全面性不足。

作者贡献：杨文怡提出研究选题方向，负责文献检索收集和整理，并撰写论文初稿；王敬鑫进行文献数据提取和提取信息整理；艾丽梅负责论文的修订；万霞负责文章的质量控制及审校，对文章整体负责；所有作者确认了论文的最终稿。

本文无利益冲突。

参考文献

[1] MENG Q，XU L，ZHANG Y，et al. Trends in access to health services and financial protection in China between 2003 and 2011：a cross-sectional study［J］. Lancet，2012，379（9818）：805-814. DOI：10.1016/s0140-6736(12)60278-5.

[2] ØDEGAARD K M，LIRHUS S S，MELBERG H O，et al. A nationwide registry study on heart failure in Norway from 2008 to 2018：variations in lookback period affect incidence estimates［J］. BMC Cardiovasc Disord，2022，22（1）：88. DOI：10.1186/s12872-022-02522-y.

[3] CZWIKLA J，JOBSKI K，SCHINK T. The impact of the lookback period and definition of confirmatory events on the identification of incident cancer cases in administrative data［J］. BMC Med Res Methodol，2017，17（1）：122. DOI：10.1186/s12874-017-0407-4.

[4] ROSENLUND M，EKSTROM N，TORNBLOM M，et al. Impact of variable look-back periods on the incidence rates of chronic diseases using real world data［J］. Pharmacoepidemiology and Drug Safety，2020，29（9）：1086-1092. DOI：10.1002/pds.5066.

[5] SCHMEDT N，KHIL L，BERGER K，et al. Incidence of multiple sclerosis in Germany：a cohort study applying different case definitions based on claims data［J］. Neuroepidemiology，2017，49（3/4）：91-98. DOI：10.1159/000481990.

[6] ABBAS S，IHLE P，KOESTER I，et al. Estimation of disease incidence in claims data dependent on the length of follow-up：a methodological approach［J］. Health Services Research，2012，47（2）：746-755. DOI：10.1111/j.1475-6773.2011.01325.x.

[7] O'BRIEN B，HARRIS T，BECKMAN J，et al. Standards for reporting qualitative research：a synthesis of recommendations［J］. Acad Med，2014，89（9）：1245-1251. DOI：10.1097/acm.0000000000000388.

[8] BRAMELD K J，HOLMAN C D，LAWRENCE D M，et al. Improved methods for estimating incidence from linked hospital morbidity data［J］. Int J Epidemiol，2003，32（4）：617-624. DOI：10.1093/ije/dyg191.

[9] BEAUDET N，COURTEAU J，SARRET P，et al. Improving the selection of true incident cases of low back pain by screening retrospective administrative data［J］. Eur J Pain，2014，18（7）：923-931. DOI：10.1002/j.1532-2149.2013.00437.x.

[10] ASGHARI S，COURTEAU J，CARPENTIER A C，et al. Optimal strategy to identify incidence of diagnostic of diabetes using administrative data［J］. BMC Med Res Methodol，2009，9：62. DOI：10.1186/1471-2288-9-62.

[11] VANASSE A，COURTEAU J，FLEURY M J，et al. Treatment prevalence and incidence of schizophrenia in Quebec using a population health services perspective：different algorithms，different estimates［J］. Soc Psychiatry Psychiatr Epidemiol，2012，47（4）：533-543. DOI：10.1007/s00127-011-0371-y.

[12] BENCHIMOL E I，GUTTMANN A，MACK D R，et al. Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario，Canada［J］. J Clin Epidemiol，2014，67（8）：887-896. DOI：10.1016/j.jclinepi.2014.02.019.

[13] LIN J，SHAU W，LAI M. Sex- and age-specific prevalence and incidence rates of sight-threatening diabetic retinopathy in Taiwan［J］. JAMA Ophthalmology，2014，132（8）：922-928. DOI：10.1001/jamaophthalmol.2014.859.

[14] SULO G，IGLAND J，VOLLSET S，et al. Effect of the lookback period's length used to identify incident acute myocardial infarction on the observed trends on incidence rates and survival：cardiovascular disease in Norway project［J］. Circ Cardiovasc Qual Outcomes，2015，8（4）：376-382. DOI：10.1161/circoutcomes.114.001703.

[15] RASSEN J，BARTELS D，SCHNEEWEISS S，et al. Measuring prevalence and incidence of chronic conditions in claims and electronic health record databases［J］. Clin Epidemiol，2019. 11：1-15. DOI：10.2147/clep.S181242.

[16] STEIN J，BLACHLEY T，MUSCH D. Identification of persons with incident ocular diseases using health care claims databases［J］. Am J Ophthalmol，2013，156（6）：1169-1175. DOI：10.1016/j.ajo.2013.06.035.

[17] CAMPLAIN R，KUCHARSKA-NEWTON A，CUTHBERTSON J，et al. Misclassification of incident hospitalized and outpatient heart failure in administrative claims data：the Atherosclerosis Risk in Communities（ARIC）study［J］. Pharmacoepidemiol Drug Saf，2017，26（4）：421-428. DOI：10.1002/pds.4162.

[18] OHLMEIER C，SAUM K，GALETZKA W，et al. Epidemiology and health care utilization of patients suffering from Huntington's disease in Germany：real world evidence based on German claims data［J］. Bmc Neurology，2019，19（1）：318. DOI：10.1186/s12883-019-1556-3.

[19] KOO B，LEE C，YANG B，et al. The incidence and prevalence of diabetes mellitus and related atherosclerotic complications in Korea：a National Health Insurance Database Study［J］.

PLoS One，2014，9（10）：e110650. DOI：10.1371/journal.pone.0110650.

[20] JÜRISSON M，VOROBJOV S，KALLIKORM R，et al. The incidence of hip fractures in Estonia，2005—2012［J］. Osteoporosis International，2015，26（1）：77-84. DOI：10.1007/s00198-014-2820-4.

[21] BUKHBINDER A，LING Y，HASAN O，et al. Risk of Alzheimer's disease following influenza vaccination：a claims-based cohort study using propensity score matching［J］. J Alzheimers Dis，2022. 88（3）：1061-1074. DOI：10.3233/jad-220361.

[22] ZHOU L，NGO L，FIELD T. Population incidence of cerebral venous thrombosis（CVT）from a large canadian cohort 2001—2017. Stroke，2022，53（s1）：61. DOI：10.1161/str.53.suppl-1.61.

[23] KATZENELLENBOGEN J，BOND-SMITH D，CUNNEEN R，et al. Contemporary Australian profile of rheumatic fever：results from multiple jurisdictions［J］. Heart Lung and Circulation，2019，28（s2）：42-43. DOI：10.1016/j.hlc.2019.05.111.

[24] OU Y，GROSSMAN D，LEE P，et al. Glaucoma，Alzheimer disease and other dementia：a longitudinal analysis［J］. Ophthalmic Epidemiology，2012，19（5）：285-292. DOI：10.3109/09286586.2011.649228.

[25] CLISSOLD B，SUNDARARAJAN V，CAMERON P，et al. Stroke incidence in Victoria，Australia-emerging improvements［J］. Front Neurol，2017，8：180. DOI：10.3389/fneur.2017.00180.

[26] AHN J，HWANG J，SEO G. Incidence and risk of developing rheumatic diseases in 19 724 patients with palindromic rheumatism in South Korea：a nationwide population-based study［J］. Joint Bone Spine，2021，88（3）：105128. DOI：10.1016/j.jbspin.2020.105128.

[27] ANDERSON K，FUHRER R，ABRAHAMOWICZ M，et al. The incidence of first-episode schizophrenia-spectrum psychosis in adolescents and young adults in montreal：an estimate from an administrative claims database［J］. Can J Psychiatry，2012，57（10）：626-633. DOI：10.1177/070674371205701007.

[28] KATZENELLENBOGEN J，SANFILIPPO F，HOBBS M，et al. Variable effects of prevalence correction of population denominators on differentials in myocardial infarction incidence：a record linkage study in Aboriginal and non-Aboriginal Western Australians［J］. J Clin Epidemiol，2011，64（6）：658-666. DOI：10.1016/j.jclinepi.2010.08.008.

[29] BYRT T，BISHOP J，CARLIN J B. Bias，prevalence and kappa［J］. J Clin Epidemiol，1993，46（5）：423-429. DOI：10.1016/0895-4356(93)90018-v.

[30] 谢莉琴，秦盼盼，高星，等. 中国城乡居民基本医疗保险制度发展历程、挑战与应对策略［J］. 中国公共卫生，2020，36（12）：1673-1676. DOI：10.11847//zgggws1126456.

[31] WU R，LI N，ERCIA A. The effects of private health insurance on universal health coverage objectives in China：a systematic literature review［J］. Int J Environ Res Public Health，2020，17（6）：1660-4601. DOI：10.3390/ijerph17062049.

（收稿日期：2023-01-05；修回日期：2023-04-18）

（本文编辑：陈俊杉）